

Metodologia de apoio à operação de sistemas elétricos de potência baseada em identificação de sistemas com aprendizado de máquina[★]

Ricardo R. Almeida^{*} Ricardo Schumacher^{**} Alexandre R. Aoki^{**}

^{*} Centro de Operações de Geração e Transmissão, Copel Geração e Transmissão SA, PR, (e-mail: ricardo.almeida@copel.com).

^{**} Departamento de engenharia elétrica, Universidade Federal do Paraná, PR (e-mail: schumacher.ric@gmail.com, araoki@gmail.com).

Abstract: Recent advances on synchro phasor measurement along with its regulation by the Operador Nacional do Sistema – ONS brought access to high resolution data with rates up to one sample per cycle. Electrical power system current, voltage, phase and frequency are collected through Phasor Measurement Units-PMUs and stored in Phasor Data Concentrators – PDCs. Such high-resolution data favours new algorithms and applications towards electrical power systems dynamic behaviour, once unattainable due to supervisory control and data acquisition system limitations. Implementations regarding real time operations help control room operators' awareness and reactions under electrical power system disturbances and malfunctions. System identification methods, together with artificial intelligence tools such machine learning, contribute to research and development on real time supervisory control. In this study a classic system identification algorithm is implemented to process dynamic electrical power system data and find a model representing its behaviour through PMU data. A stabilization diagram is applied as a post processing method and results used to build a histogram underlining electrical power systems electromechanical modes.

Resumo: Recentes avanços tecnológicos na medição sincronizada de fasores e a regulação de padrões pelo Operador Nacional do Sistema – ONS, propiciaram o acesso a dados do sistema elétrico de potência com resolução de até uma amostra por ciclo. Aplicações em operação em tempo real do sistema elétrico podem auxiliar os profissionais de sala de controle na observabilidade do comportamento dinâmico do sistema, na atuação preventiva durante perturbações e na resposta a situações de falta. Dados com alta resolução favorecem o desenvolvimento de algoritmos de identificação da dinâmica do sistema de potência, antes inviável devido às limitações de sistemas de supervisão e controle tradicionais. Técnicas e algoritmos de identificação de sistemas, associadas a ferramentas de inteligência artificial, representam um importante campo de pesquisa e desenvolvimento voltado a sistemas de supervisão e controle em tempo real. Neste trabalho, um algoritmo clássico de identificação de sistemas é utilizado para processar informações dinâmicas e estimar o modelo do sistema através dos dados coletados por PMUs. Um diagrama de estabilização é utilizado como ferramenta de pós processamento das estimativas e o resultado é apresentado em forma de histograma, destacando modos dominantes nas estimativas obtidas e caracterizando modos de oscilação eletromecânica em um determinado sistema elétrico. Resultados iniciais demonstram a capacidade de caracterização de modos oscilatórios e indicam a possibilidade de identificação de faltas.

Keywords: System identification; Machine learning; WAMS.

Palavras-chaves: Identificação de sistemas; Aprendizado de máquina; Sistemas elétricos de potência de grande porte.

1. INTRODUÇÃO

A supervisão e controle de sistemas elétricos de potência é um importante e complexo processo, cujo objetivo central é garantir a disponibilidade e qualidade no fornecimento de eletricidade. Via de regra é composto pelos equipamentos de medição e monitoramento de estado de disjuntores, seccionadoras, transformadores e linhas de transmissão, estendendo-se a relés de proteção, Unidades Terminais Remotas – UTRs, sistemas de comunicação, sistema de gerenciamento de energia – do inglês, *Energy Management System* – EMS e por fim disponibilizando os estados das variáveis discretas e contínuas ao operador de sala de controle através de estações de operação com a representação gráfica de subestações e linhas de transmissão via diagramas unifilares animados representativos do sistema elétrico em tempo real.

Em condições de normalidade, a atenção do elemento humano responsável pelo monitoramento do sistema é dividida entre a observação das telas do Sistema de Supervisão e Controle – SSC, contato telefônico com as equipes de campo, agentes de outras empresas e demais atividades correlatas, como registro de anomalias, contingências e das atividades desenvolvidas na rotina operacional.

Um sistema projetado e implantado de forma adequada terá funcionamento estável em regime permanente, suportando variações provocadas por aumento ou redução na carga, entrada ou saída de um conjunto de equipamentos. A operação bem planejada permitirá a execução de intervenções de manutenção que provoquem indisponibilidades sem afetar a qualidade e disponibilidade no fornecimento de energia elétrica. Mesmo durante perturbações, proteções e automatismos atuando corretamente irão eliminar faltas tempestiva e coordenadamente, evitando que a falha estenda-se a regiões além do necessário.

No entanto, dependendo da complexidade da perturbação, aqui entendida como variação inesperada e além dos limites nas condições de operação, ou da falta de um determinado equipamento ou conjunto de equipamentos, o reconhecimento do estado do sistema durante ou após a falta pode ser bastante prejudicado.

O desenvolvimento das ferramentas existentes e a introdução de novos sistemas, capazes de auxiliar a tarefa do operador de sala de controle é uma área de pesquisa relevante para sistemas elétricos de potência. Neste sentido, uma das mais importantes evoluções recentemente apresentadas foi a dos sistemas de medição sincrofasorial – SMSF. Embora seja uma tecnologia conhecida desde o final da década de 80, sua aplicação a sistemas de potência experimenta um aumento significativo desde o início da década passada, com a instalação de dispositivos de forma crescente em instalações de usinas e subestações.

Neste trabalho um algoritmo clássico de identificação de sistemas é utilizado para processar informações dinâmicas e estimar o modelo do sistema através dos dados coletados na PMU. Um diagrama de estabilização modificado é utilizado como ferramenta adicional de pós processamento

* Este trabalho foi financiado pelo programa de pesquisa e desenvolvimento tecnológico da Companhia Paranaense de Energia – COPEL, através do projeto PD-06491-05312019, regulado pela ANEEL.

das estimativas e o resultado apresentado em forma de histograma, de maneira a destacar os modos dominantes obtidos e caracterizando, por exemplo, a frequência e a taxa de amortecimento do modo oscilatório eletromecânico dominante de um determinado sistema elétrico.

2. MODELAGEM DE SINAL COM IDENTIFICAÇÃO DE SISTEMAS E DIAGRAMA DE ESTABILIZAÇÃO

Conforme Missout and Girard (1980), a procura por métodos diretos de medição da diferença entre ângulos de fase encontra suas primeiras referências de implementações de sucesso na década de 80, com a utilização de sinais de sincronização por satélite como por exemplo Long range navigation–Loran-C ou Geostationary Operational Environmental Satellite–GOES. Paralelamente, a evolução dos sistemas de proteção de linhas de transmissão trouxe como resultado um novo método utilizando componentes simétricas (Phadke et al., 1977) abrindo espaço para o desenvolvimento dos atuais algoritmos em uso para medição de fasores. Phadke et al. (1983) introduz a base para o que pode ser considerada a fase moderna da tecnologia de medição fasorial sincronizada. Além dos sincrofasores, frequência e taxa de variação de frequência são dados normalmente disponíveis em um SMSF.

Neste estudo, a frequência é o sinal utilizado na estimação do modelo correspondente ao sistema elétrico analisado.

2.1 Fluxograma de processamento

O processamento do sinal é feito conforme o fluxograma da Figura 1.

Dados disponíveis em um PDC de um SMSF instalado em um sistema de potência de grande porte são importados em formato .csv, emulando a etapa de coleta de dados de PMU. Os dados são então pré-processados para eliminar ruído, componentes de alta e baixa frequência indesejáveis ou sem importância para o processo, bem como para reduzir o esforço computacional das fases subsequentes. Os dados tratados são submetidos a um algoritmo de identificação de sistemas utilizando a técnica de mínimos quadrados recursivo, à qual são agregadas componentes de robustez e regularização. Os polos complexos conjugados resultantes, correspondentes à frequência e taxa de amortecimento do modo oscilatório eletromecânico do sistema de potência associado ao sinal de frequência do SMSF, são utilizados para construção de um histograma da taxa de amortecimento x frequência do modo oscilatório. A última etapa do processamento é a aplicação de um algoritmo de Detecção de Anomalias, utilizado para caracterizar a mudança no comportamento do sistema de potência, provocado por alguma variação de geração, carga ou uma perturbação sistêmica causada por falta.

2.2 Coleta e pré-processamento

A coleta dos dados é realizada de forma não automatizada, copiando arquivos salvos nos dispositivos de armazenamento do PDC, no formato CSV. A leitura dos arquivos é feita com a função *read_csv()* do pacote *pandas* (Reback et al., 2022). Dos dados disponíveis, como frequência, tensão, corrente, etc, neste estudo a informação utilizada foi o sinal de frequência do sistema.

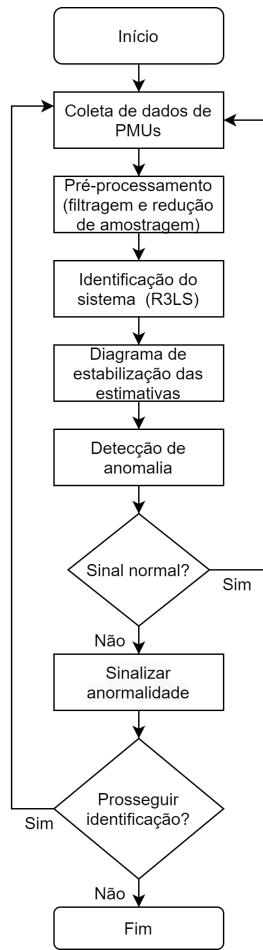


Figura 1. Fluxograma

Os dados são em seguida submetidos a filtragem e redução de amostras (reamostragem). O sinal é normalizado através da subtração pela média dos valores dos dados em análise, correspondente à eliminação da componente contínua do sinal, melhorando a relação sinal ruído. Após esta operação é aplicado um filtro passa-faixa digital do tipo Butterworth de quarta ordem. O filtro tem uma banda de passagem definida entre 0,2 e 2,5 Hz, que é a faixa de frequências de interesse para o processo em análise.

Oscilações eletromecânicas de baixa frequência, na faixa de 0,1 a 2,0 Hz, correspondem a um fenômeno inerente a sistemas elétricos de potência e suas unidades geradoras e linhas de transmissão. Tais oscilações derivam de condições dinâmicas características dos sistemas, como variações de carga, capacidade de resposta das unidades geradoras, comportamento dos sistemas de controle de tensão e frequência destas unidade durante alterações de carga e nas fontes de potência (vazão e níveis de armazenamento em reservatórios, velocidade do vento, disponibilidade de gás ou carvão, desligamentos intempestivos, etc.).

Conforme a complexidade do sistema, os modos oscilatórios podem ser decompostos de forma a analisar variações locais ou entre áreas elétricas, bem como as variáveis internas como tensão nas barras e frequência.

É de particular importância para a estabilidade de um sistema a capacidade de amortecimento das oscilações de baixa frequência, notadamente entre áreas pois tais vari-

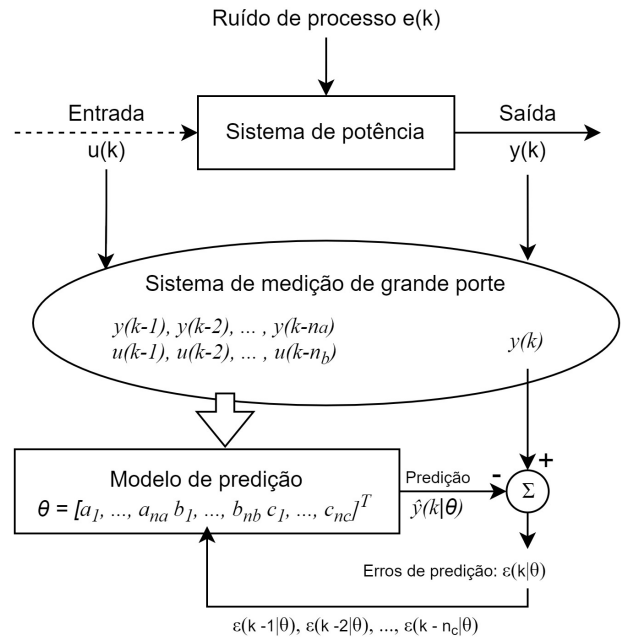


Figura 2. Modelo ARMAX de fluxo de dados

ações concorrem para perturbações que podem levar ao colapso de um sistema elétrico. Este fenômeno também é observado em redes de distribuição, com oscilações características de até 6,0 Hz conforme Kuiuava et al. (2008).

A etapa final de pré-processamento é a de redução de amostras, necessária para aproximar a amostragem ao ponto ótimo. Uma taxa amostral alta aumenta a sensibilidade de polos discretos ao ruído bem como aos polos próximos. Uma taxa de amostragem ótima encontra-se na faixa de 4 a 6 Hz para a estimação de modos eletromecânicos (Ding and Huang, 2015).

2.3 Identificação do sistema (R3LS)

Considerando o tipo de aplicação ter como propósito identificar um sistema em tempo real, a partir de dados de PMUs, a opção de implementação foi por um algoritmo recursivo com base no método dos mínimos quadrados denominado R3LS. Esta opção se justifica pela capacidade deste gênero de algoritmo armazenar os parâmetros das as estimativas com base no estado anterior e atualizá-los a partir de novas amostras.

O algoritmo foi desenvolvido em Python seguindo proposta originalmente apresentada em Zhou et al. (2008), conforme detalhamento a seguir.

Utilizando o modelo *autoregressive moving average exogenous*-ARMAX, ilustrado pela Figura 2, representado por:

$$y(k) + a_1 y(k-1) + \dots + a_{n_a} y(k-n_a) = b_1 u(k-1) + \dots + b_{n_b} u(k-n_b) + c_1 e(k-1) + \dots + c_{n_c} e(k-n_c) + e(k) \quad (1)$$

Onde $y(k)$ é o sinal de saída do sistema, $u(k)$ a entrada e $e(k)$ o ruído do processo, no instante de tempo k , com $k = n + 1, n + 2, \dots, t$ e $n = \max\{n_a, n_b, n_c\}$.

Definindo $\theta = [a_1 \dots a_{n_a} \ b_1 \dots b_{n_b} \ c_1 \dots c_{n_c}]^T$, vetor de parâmetros a ser calculado, onde n_a é a ordem do modelo

autoregressivo (AR), n_b a ordem do modelo de entradas X e n_c a ordem do modelo de média movel (MA). Calculando-se as raízes do polinômio característico $1 + a_1z^{-1} + a_2z^{-2} + \dots + a_{n_a}z^{-n_a}$, encontram-se os modos do sistema.

Definindo o erro de predição $\varepsilon(k|\theta) = y(k) - \hat{y}(k|\theta)$ e o vetor:

$$\varphi(k|\theta) = [-y(k-1) \dots - y(k-n_a) \quad u(k-1) \dots u(k-n_b) \quad \varepsilon(k-1|\theta) \dots \varepsilon(k-n_c|\theta)]^T \quad (2)$$

O modelo de predição pode ser escrito como:

$$\begin{aligned} \hat{y}(k|\theta) &= -a_1y(k-1) - \dots - a_{n_a}y(k-n_a) \\ &\quad + b_1u(k-1) + \dots + b_{n_b}u(k-n_b) \\ &\quad + c_1\varepsilon(k-1|\hat{\theta}(k-1)) \\ &\quad + \dots + c_{n_c}\varepsilon(k-n_c|\hat{\theta}(k-n_c)) \\ &= \varphi^T(k|\theta)\theta \end{aligned} \quad (3)$$

Sejam definidos

- $\bar{\theta}$, um vetor constante;
- $\Pi(t)$, uma matriz definida positiva;
- λ | $0 \ll \lambda \leq 1$ uma constante próxima da unidade, denominada fator de esquecimento;
- $\rho(\varepsilon)$ uma função de perda que define a medida dos erros.

e uma função de custo:

$$J(\theta(t)) = \frac{1}{2} (\theta(t) - \bar{\theta})^T \Pi(t) (\theta(t) - \bar{\theta}) + \sum_{k=n+1}^t \lambda^{t-k} \rho[\varepsilon(k|\theta(t))] \quad (4)$$

onde $\sum_{k=n+1}^t \lambda^{t-k} \rho[\varepsilon(k|\theta(t))]$ é utilizada para agregar "robustez" ao algoritmo, através da aplicação de uma função de redução da influência de erros de predição:

$$\rho(\varepsilon) = \begin{cases} \frac{1}{2}\varepsilon^2 & \text{se } \varepsilon \leq 3\sigma, \\ \frac{1}{2}(3\sigma)^2 & \text{se } \varepsilon > 3\sigma. \end{cases} \quad (5)$$

A variável σ é o desvio padrão de ε . Quando o erro é inferior a 3σ , ρ é uma função de mínimos quadrados comum.

A constante λ , também chamada de fator de esquecimento, é uma constante positiva e ligeiramente menor do que 1 e determina a velocidade com que os dados passados são desconsiderados. Um valor pequeno de λ melhora a capacidade do algoritmo em detectar variações nos dados. Por outro lado, a redução de dados utilizados pode elevar a variância dos parâmetros de estimação.

A matriz $\Pi(t)$ confere regularização ao algoritmo. É utilizada para agregar o conhecimento prévio que existe a respeito do modelo, a partir do vetor constante $\bar{\theta}$, que é a estimativa utilizada para o vetor de parâmetros θ . Um valor alto de $\Pi(t)$ demonstra que $\bar{\theta}$ é uma estimativa adequada para o modelo.

Com estas considerações, o valor estimado de θ , segundo o método de mínimos quadrados robusto regularizado será dado a partir da equação:

$$\hat{\theta}(t) = \arg \min_{\theta(t)} J(\theta(t)) \quad (6)$$

Reescrevendo (4) para $|\varepsilon| \leq 3\sigma$:

$$J(\theta(t)) = \frac{1}{2} (\theta(t) - \bar{\theta})^T \Pi(t) (\theta(t) - \bar{\theta}) + \frac{1}{2} [\mathbf{y}(t) - H(t)\theta(t)]^T \Lambda(t) [\mathbf{y}(t) - H(t)\theta(t)] \quad (7)$$

onde

$$\Lambda = \begin{bmatrix} \lambda^{t-n-1} & & & \\ & \lambda^{t-n-2} & & \\ & & \ddots & \\ & & & \lambda^0 \end{bmatrix} \quad (8)$$

$$\mathbf{y}(t) = \begin{bmatrix} y(n+1) \\ y(n+2) \\ \vdots \\ y(t) \end{bmatrix} \quad (9)$$

$$\mathbf{H}(t) = \begin{bmatrix} \varphi^T(n+1) \\ \varphi^T(n+2) \\ \vdots \\ \varphi^T(t) \end{bmatrix} \quad (10)$$

Para $t \geq n+1$, (7) é minimizada calculando a derivada de primeira ordem e igualando a zero:

$$\frac{\partial J(\hat{\theta}(t))}{\partial \hat{\theta}(t)} =$$

$$\begin{aligned} (\hat{\theta}(t) - \bar{\theta})^T \Pi(t) - [\mathbf{y}(t) - H(t)\hat{\theta}(t)]^T \Lambda(t) H(t) &= 0 \\ \Rightarrow \Pi(t)(\hat{\theta}(t) - \bar{\theta}) - H^T(t)\Lambda(t)[\mathbf{y}(t) - H(t)\hat{\theta}(t)] &= 0 \\ \Rightarrow \hat{\theta}(t) = \Phi^{-1}(t)s(t) \end{aligned} \quad (11)$$

onde

$$\Phi(t) \triangleq \Pi(t) + H^T(t)\Lambda(t)H(t) \quad (12)$$

$$s(t) \triangleq \Pi(t)\bar{\theta} + H^T(t)\Lambda(t)\mathbf{y}(t) \quad (13)$$

A equação (11) define o algoritmo de mínimos quadrados não recursivo, onde as matrizes $H(t)$, $\Lambda(t)$ e $\mathbf{y}(t)$ aumentam de tamanho a cada iteração, o que implica necessidade crescente de espaço de armazenamento e maior tempo de processamento na obtenção de novas estimativas.

Este problema é contornado com o a versão recursiva. As equações (12) e (13) podem ser reescritas, para $t = n+1$, como:

$$\Phi(n+1) = \Pi(n+1) + \varphi(n+1)\varphi^T(n+1) \quad (14)$$

$$s(n+1) = \Pi(n+1)\bar{\theta} + \varphi(n+1)y(n+1) \quad (15)$$

$$\hat{\theta}(n+1) = \Phi^{-1}(n+1)s(n+1) \quad (16)$$

Como $\hat{\theta}(n) = \bar{\theta}$, a estimativa inicial dos parâmetros, para $t \geq n+2$:

$$\Phi(t) = \lambda\Phi(t-1) + \Pi(t) - \lambda\Pi(t-1) + \varphi(t)\varphi^T(t) \quad (17)$$

$$s(t) = \lambda s(t-1) + (\Pi(t) - \lambda\Pi(t-1))\bar{\theta} + \varphi(t)y(t) \quad (18)$$

Utilizando (14) e (15) em (11):

$$\begin{aligned} \hat{\theta}(t) &= \Phi^{-1}(t)s(t) \\ &= \hat{\theta}(t-1) + \Phi^{-1}(t)\varphi(t)\{y(t) - \varphi^T(t)\hat{\theta}(t-1)\} \\ &\quad - \Phi^{-1}(t)(\Pi(t) - \lambda\Pi(t-1))(\hat{\theta}(t-1) - \bar{\theta}) \end{aligned} \quad (19)$$

As equações (16), (17) e (19) formam o algoritmo recursivo, em que a estimativa atual pode ser calculada apenas atualizando as anteriores usando as novas amostras de dados, o que reduz a necessidade de espaço de armazenamento e tempo de processamento continuamente crescentes, características do algoritmo não recursivo.

Como forma de reduzir custo computacional no cálculo das matrizes inversas $\Phi^{-1}(n+1)$ e Φ^{-1} em (16) e (19), aplica-se o lema de inversão de matrizes (20).

$$[A + BCD]^{-1} = A^{-1} - A^{-1}B[DA^{-1}B + C^{-1}]^{-1}DA^{-1} \quad (20)$$

Definindo:

$$P(k) \triangleq \Phi^{-1}(k), \text{ para } k = n+1, n+2, \dots, t. \quad (21)$$

Para $t = n+1$, conforme (14) e (19), deduz-se:

$$\begin{aligned} P(n+1) &= \Phi^{-1}(n+1) \\ &= [\Pi(n+1) + \varphi(n+1)\varphi^T(n+1)]^{-1} \\ &= [\Pi^{-1}(n+1) \\ &\quad - \frac{\Pi^{-1}(n+1)\varphi(n+1)\varphi^T(n+1)\Pi^{-1}(n+1)}{1 + \varphi^T(n+1)\Pi^{-1}(n+1)\varphi(n+1)}]^{-1} \end{aligned} \quad (22)$$

Conforme (22) e (15), (16) torna-se então:

$$\begin{aligned} \hat{\theta}(n+1) &= \Phi^{-1}(n+1)s(n+1) \\ &= \bar{\theta} + \frac{\Pi^{-1}(n+1)\varphi(n+1)\varphi^T(n+1)}{1 + \varphi^T(n+1)\Pi^{-1}(n+1)\varphi(n+1)} \\ &\quad \times [y(n+1) - \varphi^T(n+1)\bar{\theta}] \end{aligned} \quad (23)$$

De forma a simplificar o cálculo da matriz inversa, normalmente $\Pi(n+1)$ é uma matriz diagonal de posto completo.

Para $t \geq n+2$, a possibilidade de aplicar o lema de inversão de matrizes (20) dependerá da seleção do peso utilizado na matriz de regularização $\Pi(t)$.

Nesta aplicação a solução adotada foi um método de regularização dinâmico sugerido por Gay (1996).

Definindo:

$$\Pi(t) = \lambda\Pi(t-1) + \xi^2 p(t)p^T(t) \quad (24)$$

$$p(t) = [0 \ \dots \ 0 \ 1 \ 0 \ \dots \ 0]^T \quad (25)$$

Onde $p(t)$ é um vetor com todos elementos nulos, exceto na posição $1 + \text{mod}(t, n_{abc})$, com $n_{abc} = n_a + n_b + n_c$, em que o valor 1 é inserido. A constante ξ é um pequeno valor positivo. A partir destas definições (19) resulta:

$$\begin{aligned} \hat{\theta}(t) &= \hat{\theta}(t-1) + P(t)\varphi(t)\{y(t) - \varphi^T(t)\hat{\theta}(t-1)\} \\ &\quad - P(t)(\xi^2 p(t)p^T(t))(\hat{\theta}(t-1) - \bar{\theta}) \end{aligned} \quad (26)$$

$$\begin{aligned} P(t) &= \Phi^{-1}(t) \\ &= \{\lambda\Phi(t-1) + \xi^2 p(t)p^T(t) + \varphi(t)\varphi^T(t)\}^{-1} \end{aligned} \quad (27)$$

Aplicando novamente (20):

$$\begin{aligned} P(t) &= [\lambda\Phi(t-1) + \xi^2 p(t)p^T(t)]^{-1} \\ &\quad - \frac{[\lambda\Phi(t-1) + \xi^2 p(t)p^T(t)]^{-1} \varphi(t)\varphi^T(t)}{1 + \varphi^T(t)[\lambda\Phi(t-1) + \xi^2 p(t)p^T(t)]^{-1} \varphi(t)} \\ &\quad \cdot [\lambda\Phi(t-1) + \xi^2 p(t)p^T(t)]^{-1} \end{aligned} \quad (28)$$

E definindo $Q(t) \triangleq [\lambda\Phi(t-1) + \xi^2 p(t)p^T(t)]^{-1}$

$$P(t) = Q(t) - \frac{Q(t)\varphi(t)\varphi^T(t)Q(t)}{1 + \varphi^T(t)Q(t)\varphi(t)} \quad (29)$$

Utilizando mais uma vez (20):

$$\begin{aligned} Q(t) &\triangleq [\lambda\Phi(t-1) + \xi^2 p(t)p^T(t)]^{-1} \\ &= \frac{1}{\lambda} \left\{ P(t-1) - \frac{\xi^2 P(t-1)p(t)p^T(t)P(t-1)}{\lambda + \xi^2 p^T(t)P(t-1)p(t)} \right\} \end{aligned} \quad (30)$$

Na aplicação implementada neste trabalho foi utilizada uma constante dinâmica de regularização $\xi = 10^{-6}$.

Desenvolvendo (26), (29) e (30) o algoritmo R3LS é finalmente obtido:

$$\begin{aligned} \varepsilon(t|\hat{\theta}(t-1)) &= y(t) - \hat{y}(t|\hat{\theta}(t-1)) \\ &= y(t) - \varphi^T(t)\hat{\theta}(t-1) \end{aligned} \quad (31)$$

$$\begin{aligned} \hat{\theta}(t) &= \hat{\theta}(t-1) + P(t)\varphi(t)\rho' \left[\varepsilon(t|\hat{\theta}(t-1)) \right] \\ &\quad - P(t)(\xi^2 p(t)p^T(t))(\hat{\theta}(t-1) - \bar{\theta}) \end{aligned} \quad (32)$$

$$P(t) = Q(t) - \frac{Q(t)\varphi(t)\varphi^T(t)Q(t)\rho'' \left[\varepsilon(t|\hat{\theta}(t-1)) \right]}{1 + \varphi^T(t)Q(t)\varphi(t)\rho'' \left[\varepsilon(t|\hat{\theta}(t-1)) \right]} \quad (33)$$

$$Q(t) = \frac{1}{\lambda} \left\{ P(t-1) - \frac{\xi^2 P(t-1)p(t)p^T(t)P(t-1)}{\lambda + \xi^2 p^T(t)P(t-1)p(t)} \right\} \quad (34)$$

Onde $P(t) \in R^{n_{abc} \times n_{abc}}$ é a matriz de estados que armazena recursivamente as informações dos cálculos antes da iteração t e $\rho'(\varepsilon)$ e $\rho''(\varepsilon)$ representam a primeira e segunda derivadas da função de perda $\rho(\varepsilon)$. Considerando (5), as duas derivadas seriam:

$$\rho'(\varepsilon) = \begin{cases} \varepsilon & \text{se } \varepsilon \leq 3\sigma, \\ 0 & \text{se } \varepsilon > 3\sigma. \end{cases} \quad (35)$$

$$\rho''(\varepsilon) = \begin{cases} 1 & \text{se } \varepsilon \leq 3\sigma, \\ 0 & \text{se } \varepsilon > 3\sigma. \end{cases} \quad (36)$$

Cabe observar que quando o erro de predição ε é menor do que 3σ , o algoritmo R3LS opera como um RLS regularizado convencional. O processamento robusto só é executado quando ε for maior do que σ .

O desvio padrão σ pode ser estimado usando:

$$\hat{\sigma} \triangleq \frac{\text{median}|\varepsilon(i) - \text{median}(\varepsilon(i))|}{0,6745} \quad (37)$$

para $i = 1, 2, \dots, N_m$, e N_m o tamanho da janela de dados.

Esta é a mediana dos desvios absolutos da mediana. Esta é uma medida de dispersão similar ao desvio padrão, mas mais robusta a pontos fora da curva (Huber and Ronchetti, 2009). O divisor 0,6745 é usado pois nesta condição $\hat{\sigma}$ é aproximadamente igual à σ se N_m for grande o suficiente

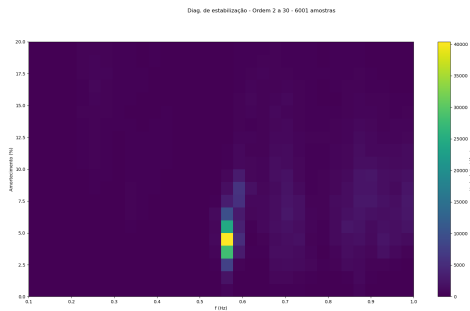


Figura 3. Diagrama de estabilização modificado

e as amostras derivam de uma função de distribuição de probabilidade normal (Kovačević et al., 1995).

2.4 Diagrama de estabilização modificado

A aplicação de diagramas de estabilização modais tem por princípio o fato dos parâmetros de frequência dos modos oscilatórios físicos terem uma fraca conexão com a ordem do modelo utilizado no processo de estimação do sistema sob análise. Dados espúrios ou indesejados, por outro lado, são afetados tanto em seus parâmetros de frequência quanto em taxa de amortecimento.

A partir desta premissa, como proposto por Kuiava et al. (2021), utilizando parâmetros calculados de frequência e taxa de amortecimento, associados às ordens dos modelos de estimação é possível avaliar o comportamento de frequência e taxa de amortecimento em função da ordem do modelo, que varia entre 2 e um valor N_m , determinado em função do número de modos oscilatórios esperados no sistema sob análise.

Um histograma de frequência x amortecimento é então montado, no qual as células correspondentes acumulam a incidência dos valores acumulados. A quantidade e tamanho das células de tal histograma está relacionada ao critério de similaridade referenciado em (38) e (39). A incidência de valores em uma determinada célula é caracterizada por um mapa de cores indicativo da contagem de amostras acumuladas em um dado par frequência x taxa de amortecimento. A Figura 3 exemplifica o formato do diagrama de estabilização proposto.

$$\left| \frac{f_i - f_{i+1}}{f_i} \right| \leq f_c = 1\% \quad (38)$$

$$\left| \frac{\xi_i - \xi_{i+1}}{\xi_i} \right| \leq \xi_c = 1\% \quad (39)$$

O procedimento de construção do diagrama de estabilização segue então os seguintes passos:

- Pré-processamento do sinal sob análise, incluindo eliminação de componentes de corrente contínua, uso de filtros passa faixa para exclusão de frequências fora da banda de interesse, melhorando a relação sinal-ruído, redução de amostras e consequente mitigação dos efeitos de perturbações de alta frequência e da concentração dos polos em torno do ponto $z - 1$ no plano complexo;
- Definição das ordens dos modelos, com valor máximo N_m , suficientemente maior do que o número de modos

eletromecânicos esperados na faixa de frequência delimitada pela etapa de pré-processamento, e um valor mínimo geralmente igual a 2;

- Construção de um histograma dos valores acumulados de frequência dos modos oscilatórios e taxas de amortecimento, em que o tamanho e quantidade de células no plano bidimensional tem por base um critério de similaridade. Faixas de valores de taxa de amortecimento (%) e frequência (Hz) delimitadas e nas células resultantes são contabilizadas as ocorrências;
- Modelos do sistema são estimados para cada amostra, de ordem N_m à menor ordem definida. Os modos estimados de todas as ordens são utilizados na composição do diagrama, de forma a compor a incidência de cada um deles no histograma.
- O histograma pode ser então representado de forma gráfica, como na Figura 3, em que no eixo x são distribuídos os valores de frequência, no eixo y são alocadas as taxas de amortecimento e a incidência em cada célula do histograma é representada por um código de cores correspondentes à contagem acumulada.

2.5 Detecção de anomalia

Validada a utilização do pré-processamento através dos métodos de identificação de sistemas e diagrama de estabilização em um sistema de referência, o passo seguinte foi avaliar a resposta dos algoritmos quando submetidos a dados ambiente coletados de PMUs instaladas em um sistema real, no caso prático, utilizando os dados de PMUs da Companhia Paranaense de Energia (COPEL).

Foram utilizados dois algoritmos disponíveis na plataforma aberta de aplicações *scikitlearn: support vector machines-SVM* e *local outlier factor-LOF*. O primeiro faz parte do módulo SVM implementado no objeto *svm.OneClassVM*, conforme Schölkopf et al. (2001). Faz uso de um *kernel* correspondente à probabilidade de encontrar uma nova observação pertencente à distribuição de treinamento, fora da fronteira estabelecida. O algoritmo LOF calcula um fator que reflete o grau de anormalidade das observações. Mede a densidade local de desvios de um dado com relação à vizinhança. A meta é identificar amostras que tem uma densidade substancialmente maior do que a da vizinhança (Breunig et al., 2000).

O princípio destes dois programas é o de identificar se uma nova amostra de um determinado conjunto de dados é similar à distribuição observada em um conjunto de dados utilizados em uma etapa de treinamento. De forma geral, é um processo de delimitação de uma fronteira da distribuição de um conjunto inicial de observações e a partir desta estimar se a distribuição de novas entradas encontra-se dentro dos limites, ou seja, se fazem parte do conjunto de dados das observações iniciais ou de treinamento.

3. CASOS DE APLICAÇÃO

Foram considerados, a fim de avaliar a proposta de tratamento dos dados de processo, dois casos de aplicação: o sistema de 39 barras de New England e o sistema Copel de PMUs.

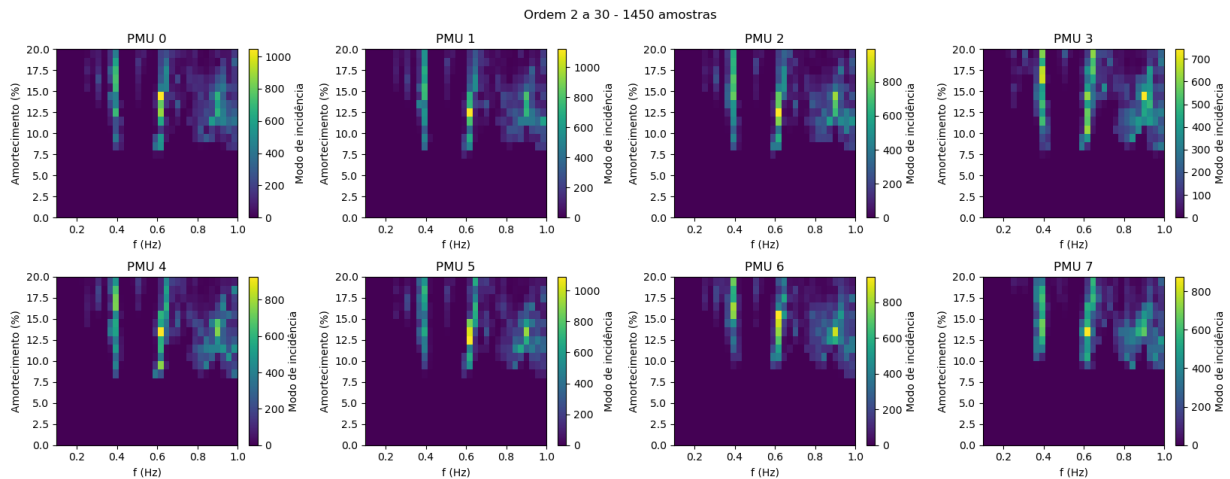


Figura 4. Sinal de frequência - 5 minutos



Figura 5. Detecção de anomalias BTA 10 minutos

Figura 7. Detecção de anomalias MGA



Figura 6. Detecção de anomalias BTA 10 minutos

Figura 8. Detecção de anomalias MGA

3.1 Sistema de 39 barras de New England

Uma primeira avaliação foi realizada com dados de frequência gerados pela simulação do sistema de 39 barras de New England, através do programa computacional PacDyn do Centro de Pesquisas de Energia Elétrica-Cepel.

A Figura 3 apresenta o resultado do processamento dos dados do modelo de referência. Neste teste foram processados os dados do sinal de frequência, calculando-se as estimativas para modelos de ordem 2 a 30. Para cada ordem foram extraídos os valores de frequência e taxa de amortecimento, acumulados de acordo com os critérios de admissão abordados em (38) e (39). Conforme a distribuição dos modos de incidência no histograma em duas dimensões, verifica-se uma concentração na célula destacada em amarelo, para frequência e taxa de amortecimento, respectivamente 0,5632 Hz e 4,46 %.

3.2 SMSF real

Na análise dos dados de um sistema elétrico de potência de grande porte real, foram utilizados os registros do sistema de medição sincrofásorial da Copel Geração e Transmissão. A Figura 4 apresenta o resultado do teste com dados

de cinco minutos do sinal de frequência de oito PMUs, com dados coletados às 18h49min do dia 02 de abril de 2021 no SMSF da Copel, nas subestações: Foz do Iguaçu Norte 230 kV (PMU 0), Cascavel Oeste 525 kV (PMU 1), Cascavel Oeste 230 kV (PMU 2), Londrina 230 kV (PMU 3), Maringá 230 kV (PMU 4), Bateias 525 kV (PMU 5), LT Itatiba (PMU 6) e Fernão Dias (PMU 7).

3.3 Detecção de anomalia

Foram utilizados sinais de frequência de uma PMU instalada na subestação Bateias 500 kV e de um evento de subfrequência coletado na subestação Maringá 230 kV e em cada caso comparados os diagramas de estabilização em três pontos do sinal de frequência. O primeiro diagrama em cada caso definido como conjunto de treinamento e a partir do segundo e terceiro conjuntos de dados obtida uma contagem de outliers. O objetivo é verificar, no primeiro caso, o comportamento do algoritmo de detecção de anomalias com um sinal ambiente sem perturbações a ele associadas e no segundo caso, uma perturbação no sistema. Em cada diagrama são registradas a quantidade de anomalias entre dois grupos de amostras, contabilizadas em relação ao total de células do histograma 32x20, no total

de 640. Os resultados são os apresentados nas Figuras 5 a 8. A contagem de *ouliers* na Figura 5, em 108/640, indica a similaridade entre os dados de um sinal de frequência em regime normal, ratificado pela contagem de 80/640 na Figura 6. Na análise do sinal da subestação MGA, observa-se na Figura 7 o sinal em uma etapa anterior à perturbação, com uma contagem de *ouliers* de 66/640, e na Figura 8 o sinal de frequência na mesma subestação segundos após uma perturbação de frequência.

4. ANÁLISE DOS RESULTADOS E CONCLUSÕES

Em relação ao sistema de referência os ensaios executados resultaram em valores similares aos obtidos por Kuiuava et al. (2021) em proposta de tratamento similar, porém com outra técnica de identificação de sistemas. Nos testes realizados com uma janela de cinco minutos, os modos eletromecânicos estão consistentes entre sub-regiões elétricas no estado do Paraná, tendo em vista os histogramas da Figura 4, com frequências aproximadas de 0,4, 0,6 e 0,9 Hz. Através da análise de duas janelas de tamanhos diferentes e em momentos distintos os modos eletromecânicos identificados através do método proposto estão de acordo com as referências existentes a respeito do sistema interligado nacional-SIN conforme a análise de Hauser and Taranto (2022). A Tabela 1 resume as taxas de amortecimento e frequência dos modos encontradas nesta análise para o período de 2020 a 2024.

Tabela 1. MODOS INTERÁREAS DO SIN ENTRE 2020-2024

Ano	NS		NNE		SSE	
	Amort.	f(Hz)	Amort.	f(Hz)	Amort.	f(Hz)
2020	9,8 %	0,46	8,5 %	0,76	16,0 %	0,79
2021	9,3 %	0,47	7,4 %	0,72	13,3 %	0,86
2022	8,8 %	0,50	8,9 %	0,75	14,8 %	0,75
2023	11,5 %	0,49	8,6 %	0,75	17,4 %	0,73
2024	5,8 %	0,46	9,2 %	0,73	16,2 %	0,74

O método proposto demonstrou ser capaz de identificar modos oscilatórios, compatíveis com a faixa esperada para a região geoeletrica. Os testes de detecção de anomalias precisam ser aprofundados, tanto quanto à capacidade de distinguir dados ambiente de perturbações, quanto à variedade de faltas avaliadas, tendo em vista a pouca disponibilidade de eventos reais no momento dos testes. Os ensaios preliminares, ainda que restritos em quantidade, são promissores quanto à capacidade de identificar perturbações.

REFERÊNCIAS

Breunig, M.M., Kriegel, H.P., Ng, R.T., and Sander, J. (2000). Lof: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 93–104.

Ding, S. and Huang, H. (2015). Improvements for electromechanical oscillation mode estimation via subspace identification methods. *EURASIP Journal on Advances in Signal Processing*, 2015, 59. doi:10.1186/s13634-015-0218-x.

Gay, S.L. (1996). Dynamically regularized fast rls with application to echo cancellation. In *1996 IEEE International Conference on Acoustics, Speech, and Signal*

Processing Conference Proceedings, volume 2, 957–960. IEEE.

Hauser, D. and Taranto, G.N. (2022). Analysis of interarea modes of the brazilian power system using the dominant pole spectrum eigensolver algorithm. *Journal of Control, Automation and Electrical Systems*, 1–8.

Huber, P. and Ronchetti, E. (2009). *Robust Statistics*, volume 30. John Wiley & Sons, Ltd. doi:10.1002/9780470434697.

Kovačević, B., Milosavljević, M., and Veinović, M. (1995). Robust recursive ar speech analysis. *Signal Processing*, 44(2), 125–138. doi:https://doi.org/10.1016/0165-1684(95)00019-A.

Kuiuava, R., Ramos, R.A., de Oliveira, R., and Bretas, N.G. (2008). An analysis of the potential impacts of electromechanical oscillations on the stability and power quality of distributed generation systems. In *2008 IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, 1–7. doi:10.1109/PES.2008.4596529.

Kuiuava, R., dos Passos, E.Z., da Costa Oliveira, G.H., and Schumacher, R. (2021). Separating low-frequency electromechanical modes from spurious modes in ambient data analysis by means of modal stabilization diagrams. *Electric Power Systems Research*, 201, 107502.

Missout, G. and Girard, P. (1980). Measurement of bus voltage angle between montreal and sept-iles. *IEEE Transactions on Power Apparatus and Systems*, PAS-99(2), 536–539. doi:10.1109/TPAS.1980.319698.

Phadke, A.G., Thorp, J.S., and Adamiak, M.G. (1983). A new measurement technique for tracking voltage phasors, local system frequency, and rate of change of frequency. *IEEE Transactions on Power Apparatus and Systems*, PAS-102(5), 1025–1038. doi:10.1109/TPAS.1983.318043.

Phadke, A., Ibrahim, M., and Hlibka, T. (1977). Fundamental basis for distance relaying with symmetrical components. *IEEE Transactions on Power Apparatus and Systems*, 96(2), 635–646. doi:10.1109/T-PAS.1977.32375.

Reback, J., McKinney, W., den Bossche, J.V., Augspurger, T., Cloud, P., Hawkins, S., Roeschke, M., Klein, A., Hoefler, P., Petersen, T., Tratner, J., She, C., Ayd, W., Naveh, S., Garcia, M., Darbyshire, J., Schendel, J., Hayden, A., Shadrach, R., Saxton, D., Gorelli, M.E., Li, F., Zeitlin, M., Jancauskas, V., McMaster, A., Battiston, P., and Seabold, S. (2022). pandas-dev/pandas: Pandas 1.4.0rc0. doi:10.5281/zenodo.5824773. URL <https://doi.org/10.5281/zenodo.5824773>.

Schölkopf, B., Platt, J., Shawe-Taylor, J., Smola, A., and Williamson, R. (2001). Estimating support of a high-dimensional distribution. *Neural Computation*, 13, 1443–1471. doi:10.1162/089976601750264965.

Zhou, N., Trudnowski, D.J., Pierre, J.W., and Mittelstadt, W.A. (2008). Electromechanical mode online estimation using regularized robust RLS methods. *IEEE Transactions on Power Systems*. doi:10.1109/TPWRS.2008.2002173.